Assignment 4 64060

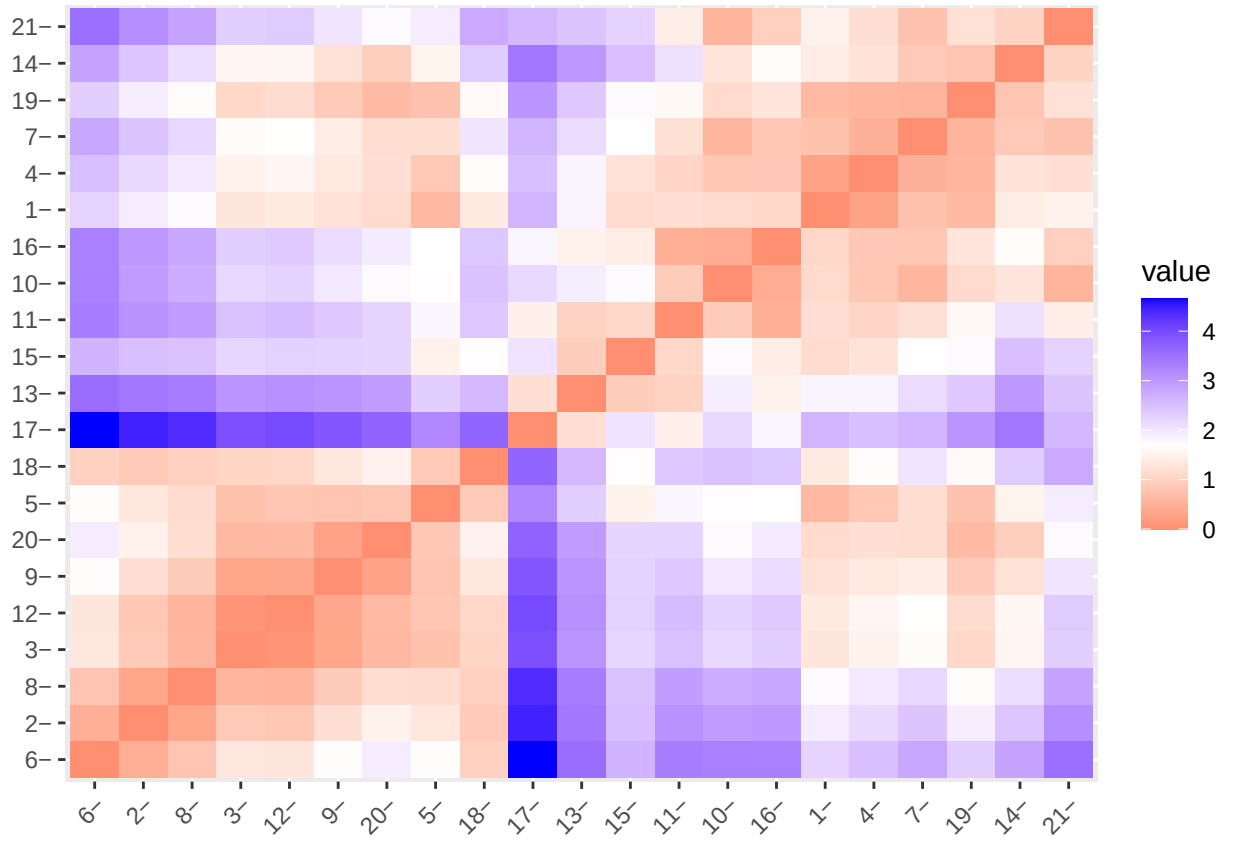
Nate Cvelbar

2023-10-29

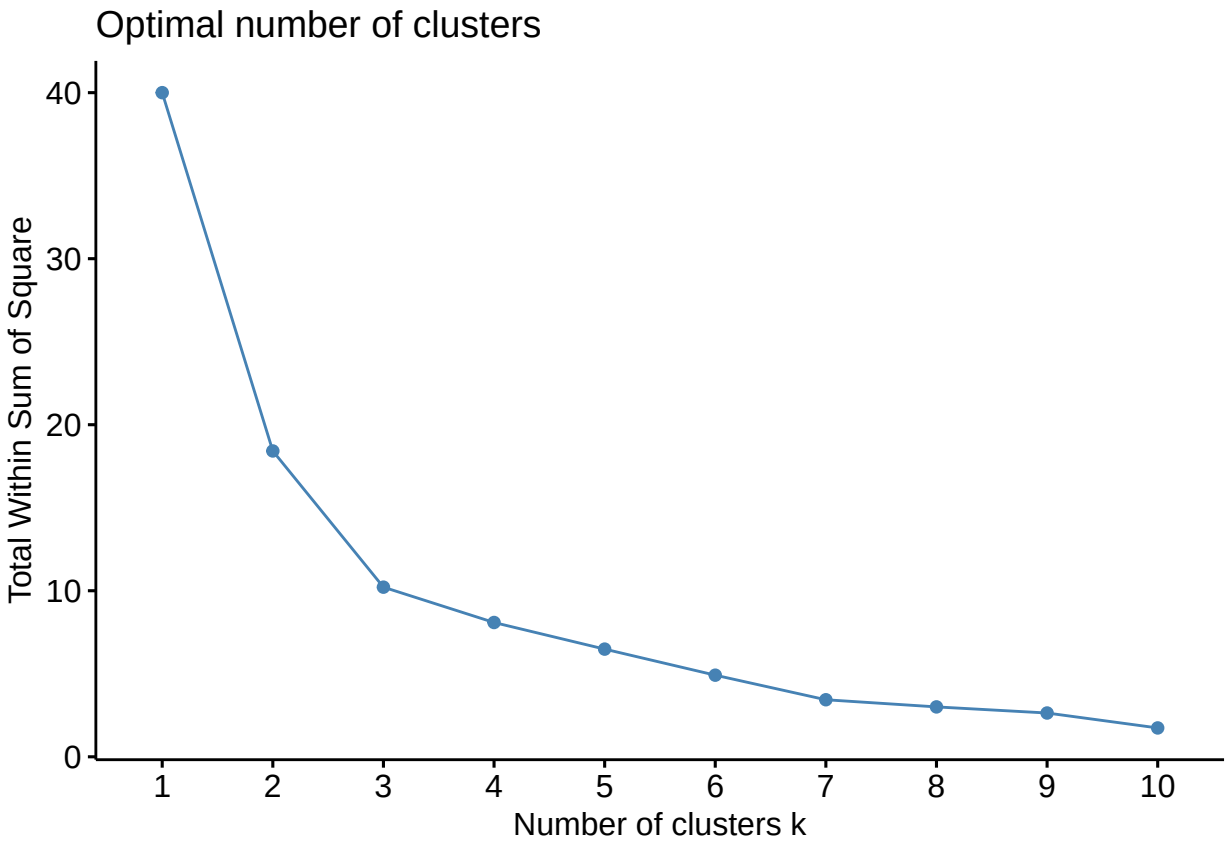
```
#install.packages("caret")
library(caret)

#install.packages("ISLR") # only install if needed
library(ISLR)

#install.packages("tidyverse") # only install if needed
library(tidyverse)
#install.packages("cluster") # only install if needed
library(cluster)
#install.packages("factoextra") # only install if needed
library(factoextra)
#install.packages("NbClust") # only install if needed
library(NbClust)
```

```
#Assignment 4
#Nate Cvelbar
#BA-64060

#File taken online from course Assignment 4
#Loading the dataset
pharm=read.csv('C:/Users/Owner/Documents/Pharmaceuticals.csv')
```

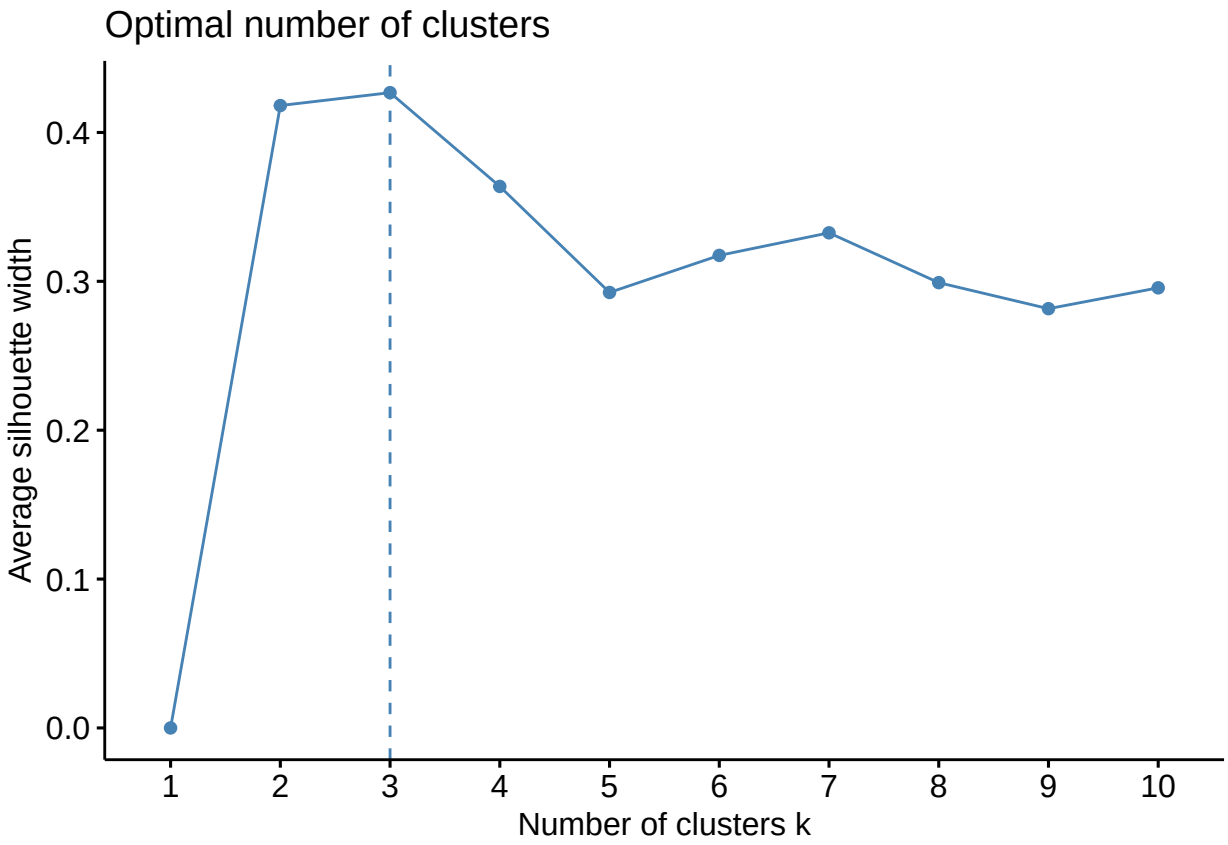
```
set.seed(111) #Set seed
#scale the data
pharmN=pharm[,c(3,11)]
df=scale(pharmN)
distance=get_dist(df)
fviz_dist(distance)
```




```
#Determine how many clusters to use. I used the elbow and silhouette methods in order to have a most ac

fviz_nbclust(df, kmeans, method = "wss")
```




```
fviz_nbclust(df, kmeans, method = "silhouette")
```




```
K=3
```

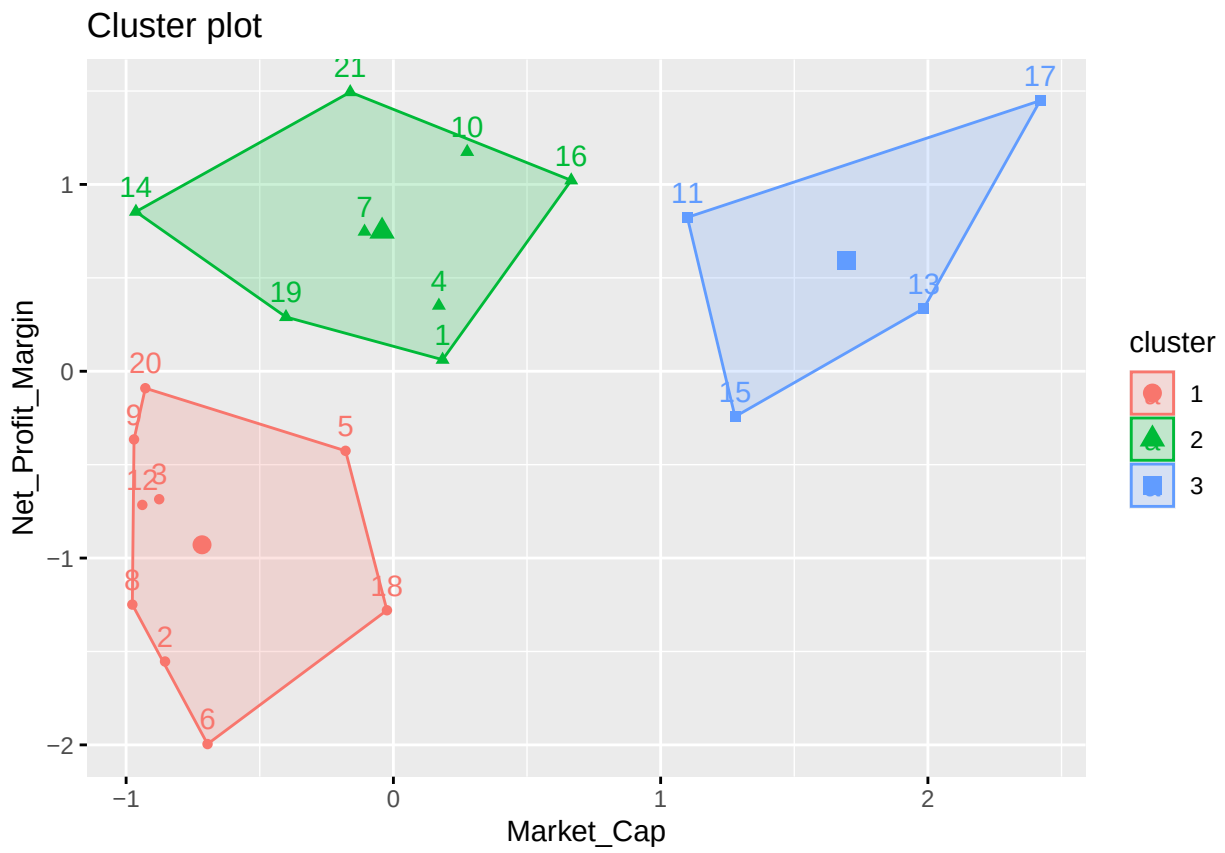
```
#Start kmeans
k3<- kmeans(df,centers = 3, nstart = 25)
k3$centers
```

```
##   Market_Cap Net_Profit_Margin
## 1 -0.7159913        -0.9288621
## 2 -0.0423004         0.7493486
## 3  1.6955811         0.5912425
```

```
k3$size
```

```
## [1] 9 8 4
```

```
fviz_cluster(k3, data = df) # Visualize the output so we can see the clusters
```




```
#The circle cluster represents small companies that are doing better than average financially. I would
#The square cluster represents small companies who are doing poorly. I would name this cluster 'Dire St
#The triangle cluster represents large companies that are doing well. I would name this cluster 'Indust
#Unfortunateley, I cannot see any relation between the clusters and the values in the last few columns
```